

# **Codebook for Replication of ‘Capital Punishment’: Bargaining and the Geography of Civil War**

Charles Butcher

[charles.butcher@otago.ac.nz](mailto:charles.butcher@otago.ac.nz)

This codebook is a guide to the variables and method that can be used to replicate the analyses and online appendix in ‘Capital Punishment’. There are seven files included in these replication materials, in addition to this codebook:

- (1) **butcher2014replication.dta**: replication file for conflict-year analysis based on the UCDP/PRIO Armed Conflict dataset from 1975-2011
- (2) **butcher2014crosssection.dta**: replication file for conflict-episode analysis based on the UCDP/PRIO Armed Conflict dataset from 1975-2011.
- (3) **butcher2014monthly.dta**: replication data file for conflict month analysis based on the UCDP/PRIO Armed Conflict Dataset and the Georeferenced Event Dataset for Africa from 1989-2010.
- (4) **butcher2014replication.do**: command file to replicate the analysis in STATA (performed in STATA version 11).
- (5) **capitaldistancedatasetv1.0.xls**: data file with measurements on minimum distance of conflict from the capital, from 1946-2011.
- (6) **butcher2014log.scml**: log file displaying the full statistical output of the analyses in the JPR article.
- (7) **Onlineappendix.docx**: Appendix showing the results of additional robustness tests mentioned, but not shown in the article. These results are also shown in the log file.

The tables and figures in the JPR article can be replicated by running the command file with these replication materials (**butcher2014replication.do**). Below is a description of the variables called in the script file. There are also 2 new variables introduced in the paper, the minimum distance of fighting from the capital city in a conflict year, and fractionalization and polarization measures. The coding of these variables is discussed in more detail below.

**Conflict-Year                      Data                      File                      Variable                      Descriptions**  
**(butcher2014replication.dta)**

Variable	Description
confid	UCDP/PRIO conflict ID (Gleditsch et al 2002)
location	Location of conflict (UCDP/PRIO ACD)
sidea	Name of side A in conflict (UCDP/PRIO ACD)
sidea2nd	Name of secondary supporter(s) for side A (UCDP/PRIO ACD)
sideb	Name of side B in conflict (UCDP/PRIO ACD)
sideb2nd	Name of secondary supporter(s) for side B (UCDP/PRIO ACD)
year	Year
coup	Conflict identified as a coup in Powell and Thyne

	(2011), 1 = yes, 0 = no
divided	Rebels internally divided, 1 = yes, 0 = no. Rebels are internally divided if they (a) have low central control in Cunningham, Gleditsch & Salehyan (2013) or are an alliance of pre-existing rebel groups in the UCDP/PRIO Conflict Actor Dataset. 1 = internally divided rebel group, 0 = not an internally divided rebel group.
capdist	Minimum distance of conflict from the capital city in the conflict year (in km). See below for more details.
incomp	Incompatibility, 1 = territory, 2 = government (Gleditsch et al, 2002)
gwno	Country number corresponding to Gleditsch and Ward (1999)
fractionalization	Fractionalization score for the conflict year. Based on the number and relative size of the government and rebel groups. Calculated using the same formula as ethnic fractionalization (Montalvo & Reynal-Querol 2005). Higher numbers indicate greater fractionalization (i.e more, evenly distributed conflict actors).
polarization	Polarization score for the conflict year. Based on the number and relative size of the government and rebel groups. Calculated using the same formula as ethnic polarization (Montalvo & Reynal-Querol 2005). Measures how close the distribution of capabilities is to bipolarity. Higher numbers indicate greater polarization (i.e the conflict is closer to a 50:50 distribution of capabilities).
bipolar	Dichotomous indicator of a 'bipolar conflict', 1= bipolar conflict, 0= not a bipolar conflict. Coded where the government controls less than 70% of capabilities and one rebel actor controls more than 30%
multipolar	Dichotomous indicator of a 'multipolar conflict', 1=multipolar conflict, 0=not a multipolar conflict. Coded where the government controls less than 70% of capabilities and two or more rebel actors control 10% of capabilities each.
in10k	Dichotomous indicator of whether conflict occurred within 10km of the capital city in the conflict year. Based on capdist. 1= fighting within 10 km of the capital, 0=fighting outside 10km of the capital.
lnrgdpchimp	Logged GDP per capita score from Gleditsch (2002) and the Penn World Tables version 7.1 (Heston,

	Summers and Aten 2012)
lnimrlag1	Logged infant mortality rate from the World Bank (2014), lagged 1 year
lnmilper	Logged military personnel from National Material Capabilities Data (Singer 1987)
lnarea	Logged land area of the country from World Bank (2014)
rebstrord	Ordinal measure of the strength of the strongest rebel group in the conflict-year, from Cunningham, Gledistch & Salehyan (2013).
fractionalizationlag1	Fractionalization, lagged 1 year
polarizationlag1	Polarization, lagged 1 year
in10klag1	in10k, lagged 1 year
mt	Proportion of mountainous terrain in the conflict zone, from Buhaug, Gates and Lujala (2009)
frst	Proportion of forested terrain in the conflict zone, from Buhaug, Gates and Lujala (2009)
alldrugs	Dichotomous indicator of drug production the conflict zone, from Buhaug, Gates and Lujala (2009).
anoc_ons	Anocracy at conflict onset, from Buhaug, Gates and Lujala (2009)
democ_on	Democracy at conflict onset, from Buhaug, Gates and Lujala (2009)
hydrod	Dichotomous indicator of hydrocarbon production the conflict zone, from Buhaug, Gates and Lujala (2009)
allgemsp	Dichotomous indicator of gem production the conflict zone, from Buhaug, Gates and Lujala (2009)
warmonths	War months elapsed since the beginning of the conflict.
warmonths2	War months elapsed since the beginning of the conflict, squared (Carter and Signorino, 2010)
warmonths3	War months elapsed since the beginning of the conflict, cubed (Carter and Signorino, 2010)
lncapdist	Logged capdist
nwstate	New state status, based on Fearon and Latin (2003), taken from Buhaug, Gates and Lujala (2009)
oil1	Oil exporter, based on Fearon and Latin (2003), taken from Buhaug, Gates and Lujala (2009)
rpe_agrilag1	Relative Extractive Capacity, Relative Political Capacity Dataset, (Arbetman-Rabinowitz et al 2013), lagged 1 year.
rpr_worklag1	Relative Political Reach, Relative Political Capacity Dataset (Arbetman-Rabinowitz et al 2013), lagged 1 year.

lncapdistlag1	Logged capdist measure, lagged 1 year.
europe	Dummy variable for whether county was in Europe, based on UCDP/PRIO indicators
mideast	Dummy variable for whether county was in the Middle East, based on UCDP/PRIO indicators
seasia	Dummy variable for whether county was in Southeast Asia, based on UCDP/PRIO indicators
africa	Dummy variable for whether county was in Africa, based on UCDP/PRIO indicators

**Conflict-Episode      Data      File      Variable      Descriptions**  
**(butcher2014crosssection.dta)**

Variable	Description
epid	Unique Episode ID based on UCDP/PRIO conflict ID and the start-date of the conflict (startdate2).
gwno	country number corresponding to Gleditsch and Ward (1999)
lnarea	logged land area of country experiencing conflict, from World Bank (2014)
fractionalization	Average fractionalization score over the conflict episode
polarization	Average polarization score over the duration of the conflict episode
in10k	Proportion of conflict-year that experienced fighting within 10km of the capital city
coup	Average number of coups in conflict episode (based on Powell and Thyne 2011, as above).
lnrgdpchimp	Average, logged GDP per capita value over the conflict-episode, from Heston, Summers and Aten (2012)
incomp	Incompatibility, 1 = territory, 2 = government. From UCDP/PRIO Armed Conflict Dataset (Gleditsch et al 2002).
divided	Dummy variable indicating whether a an internally divided rebel group was involved in the conflict episode.
lnmilper	Highest number of military personnel involved in the conflict episode, from the National Material Capabilities Dataset, version 4.0 (Singer, 1987).
warmonths	Length of conflict episode in months
warmonths2	Length of conflict episode in months, squared (Carter & Signorino, 2010).
warmonths3	Length of conflict episode in months, cubed (Carter

	& Signorino, 2010).
year	year
Inmeancapdist	Average distance of fighting from the capital over the conflict episode.

### Conflict-Month Data File Variable Descriptions (butcher2014monthly.dta)

Variable	Description
year	Year
id	UCDP/PRIO conflict ID
month	Month of observation, 1=January 12=December
sidea	Name of side A in conflict
sidea2nd	Name of secondary supporter(s) for side A
sideb	Name of side B in conflict
sideb2nd	Name of secondary supporter(s) for side B
rebstord	Ordinal measure of the strength of the strongest rebel group in the conflict-month, from Cunningham, Gleditsch & Salehyan (2013).
fractionalization2	Fractionalization score for the conflict month. Based on the number and relative size of rebel groups active in that month. Monthly entry and exit dates for rebel groups taken from Cunningham, Gleditsch & Salehyan (2013). Calculated using the same formula as ethnic fractionalization (Montalvo & Reynal-Querol 2005). Higher numbers indicate greater fractionalization (i.e more, evenly distributed conflict actors). Please note that this measure also includes formally organized non-state actors involved in non-state conflicts in the given conflict month, based on Sundberg, Eck & Kreutz (2012).
polarization2	Polarization score for the conflict month. Based on the number and relative size of rebel groups active in that month. Monthly entry and exit dates for rebel groups taken from Cunningham, Gleditsch & Salehyan (2013). Calculated using the same formula as ethnic polarization (Montalvo & Reynal-Querol 2005). Measures how close the distribution of capabilities is to bipolarity. A score of '1' is a bipolar conflict. Please note that this measure also includes formally organized non-state actors involved in non-state conflicts in the given conflict month, based on Sundberg, Eck & Kreutz (2012).
bipolar	Dichotomous indicator of a 'bipolar conflict'. Coded where the government controls less than 70% of

	capabilities and a rebel actor controls more than 30% in the conflict month.
multipolar	Dichotomous indicator of a ‘multipolar conflict’. Coded where the government controls less than 70% of capabilities and two or more rebel actors control 10% of capabilities each in the conflict month.
incomp	Incompatibility, 1 = territory, 2 = government. From UCDP/PRIO Armed Conflict Dataset (Gleditsch et al 2002).
location	Location of conflict
gwnoa	Country number corresponding to Gleditsch and Ward (1999) for side A
coup	Conflict identified as a coup in Powell and Thyne (2011), 1 = yes, 0 = no
fractionalization2lag1	Fractionalization2 score lagged by 1 month
polarization2lag1	Polarization2 score lagged by one month
fractionalizationlag1	Fractionalization score that does not include formally organized actors in non-state conflicts in the calculation, lagged by 1 month
polarizationlag1	Fractionalization score that does not include formally organized actors in non-state conflicts in the calculation, lagged by 1 month
lnarea	Logged land area of country experiencing conflict, from World Bank (2014)
warmonths	Number of months elapsed since the start-date of the conflict episode (startdate2).
epid	Unique conflict episode ID based on UCDP/PRIO conflict ID and the start-date of the episode (startdate2).
in10k20	Dummy variable indicating whether a battle occurred within 10km of the capital city in which 20 or more battle-related deaths were observed. Data on violent conflict events taken from the GED (Sundberg and Melander 2013). 1= a battle within 10km of the capital resulting in 20 deaths occurred, 0 = no such battle occurred.
in25k20	Dummy variable indicating whether a battle occurred within 25km of the capital city in which 20 or more battle-related deaths were observed. Data on violent conflict events taken from the GED (Sundberg and Melander 2013). 1= a battle within 25km of the capital resulting in 20 deaths occurred, 0 = no such battle occurred.
in10kmean	Dummy variable indicating whether the average distance of fighting from the capital was within 10km

	of the capital, based on all violent conflict events related to the relevant conflict ID in the conflict month. Data on violent conflict events taken from the GED (Sundberg and Melander 2013). 1 = the average distance of conflict events was less than 10km from the capital, 0 = the average distance of conflict events was greater than 10km from the capital.
in25kmean	Dummy variable indicating whether the average distance of fighting from the capital was within 25km of the capital, based on all violent conflict events related to the relevant conflict ID in the conflict month. Data on violent conflict events taken from the GED (Sundberg and Melander 2013). 1 = the average distance of conflict events was less than 25km from the capital, 0 = the average distance of conflict events was greater than 25km from the capital.
dividedlag1	Internally divided rebel groups variable, lagged by 1 month. Based on whether there was an internally divided rebel group active in the conflict month. Rebels are internally divided if they (a) have low central control in Cunningham, Gleditsch & Salehyan (2013) or are an alliance of pre-existing rebel groups in the UCDP/PRIO Conflict Actor Dataset. Monthly entry and exit dates for rebel groups taken from Cunningham, Gleditsch & Salehyan (2013). 1 = internally divided rebel group, 0 = not an internally divided rebel group.
lnmilperlag1	Logged military personnel from National Material Capabilities Data version 4.0 (Singer 1987), lagged 1 month
lngdpchimlag1	Logged GDP per capita score from Gleditsch (2002) and the Penn World Tables version 7.1 (Heston, Summers and Aten 2012), lagged 1 month
multipolar2lag1	‘Multipolar’ variable, discussed above, lagged by 1 month
bipolar2lag1	‘Bipolar’ variable, discussed above, lagged by 1 month

### **New Variables    Minimum Conflict Distance from the Capital City**

The JPR article uses new data on the minimum distance of fighting from the capital city. These data were collected for every civil war year in the



UCDP/PRIO armed conflict dataset for 1946-2011. The results in the paper only report the findings with the data from 1975-2011, but the full dataset is attached as a part of this replication file (capitaldistancedatasetv1.0.xls).

The ‘capdist’ variable measures the distance from the nearest battle point between the active parties to the conflict in that year, and the central business district of the capital city. A ‘battle point’ is defined as the location where 20 or more deaths were sustained in fighting between the security forces of the government and the rebel groups(s) active in the conflict year. The threshold of 20 deaths was designed to avoid coding massacres and terrorist attacks as indicators of insurgent rebel strength, of which they often are not. Thus, for example, suicide bombings do not count as a ‘battle point’ even if there are a large number of people killed in the attack. This threshold and conceptualization is consistent with the aims of the hypotheses in the JPR article that seek to examine, specifically, the closest location of insurgent strength to the capital, not the ability of rebel groups to infiltrate small numbers of insurgents and attack the capital in this way. In cases where there was very little fighting, or a battle that resulted in 20 or more deaths did not occur, information on the location of rebel ‘bases’ or their main area of operations in that year were used as ‘battle points’. This occurred only in low-level insurgencies that were often fought very far from the capital (see below with regards to the error structure of this variable). The ‘in10k’ variable is based on ‘capdist’ and measures whether a battle resulting in more than 20 deaths between the security forces and armed insurgents occurred within 10 km of the capital in that conflict year. Dates for the onset of fighting within 10 km of the capital will be released with the next version of the data, along with an update to 2013.

This variable was coded using three main sources: *Keesings Contemporary Archives/Record of World Events*, *Proquest Historical Newspapers*, and *Factiva Searches*. Secondary sources were used, on occasions where media sources did not report any conflict events. I have also used narratives from the UCDP Conflict Encyclopedia and the data from Buhaug and Gates (2002) *The Geography of Civil War*, the *Armed Conflict Location and Event Dataset* (Clionadh et al 2010) and the *Georeferenced Event Dataset* (Sundberg & Melander 2013), in addition to Hallberg’s (2012) Conflict Site Dataset. The data for 1989-2011 was coded primarily with newswires from Factiva searches, especially *Reuters* news and *The Associated Press*. Data from 1980-1988 are taken primarily from Proquest historical newspapers, some Reuters and AP newswires and, *Keesings Contemporary Archives/Record of World Events*. Data from 1946-1979 is sourced primarily from *Keesings Contemporary Archives/Record of World Events* and *Proquest Historical Newspapers*, especially the *New York Times*, *The Guardian*, and the *Wall Street Journal*.

Users should assume that there is error in the measurements. However, these errors are likely to be systematically related to the distance of fighting

form the capital. Media sources report more events, and report them more accurately when fighting is near to the capital (Kalyvas 2004). Estimates of fighting close to the capital are likely to be quite accurate, especially the fighting within 10 km of the capital dichotomous variable. Point estimates of conflict far from the capital may involve considerable error, especially for separatist conflicts that are low-intensity and poorly reported. One example is the conflict between Muslim insurgents in the Arakan region of Myanmar in the early 1950s through to the mid 1970s. While it is fairly clear that much of the activity was located near the border between Bangladesh/East Pakistan and Myanmar, near the Buthidaung region and Maungdaw in Arakan state (Moshe 2002), the exact battle locations are very difficult to ascertain. In this case the measurement is taken from these areas (i.e from Buthidaung to Rangoon) while the actual fighting may have been further or closer to Rangoon than this. It is for this reason that users are recommended to use a logged version of the ‘capdist’ variables in their analyses. It is also for this reason that the findings in the JPR paper are tested in the sample of high intensity conflicts where battle locations are more likely to be accurately reported, and in the sample of governmental conflicts, where a similar principle may apply.

The distance of conflict from the capital has been used in this paper as a dependent variable – i.e something to be explained. But it is likely that the ‘capdist’ variable is also a time sensitive indicator of the strength of insurgent forces and general government collapse, and researchers may find the variable useful for other projects. There are other indicators of rebel strength, including troop numbers and the relative strength variable from Cunningham, Gleditsch and Salehyan (2013). Annualized data are difficult to collect on these variables however, because measures of troop size, etc, are generally not available year-on-year, even in the present media environment. The distance of conflict from the capital may be an observable indicator of insurgent capability. Indeed, in low information environments the geography of civil war may be one of the primary ways in which outside actors assess the probability of government collapse (Greig 2013).

### **Conflict Actor Fractionalization and Polarization**

The main independent variable in the study is conflict actor ‘fractionalization’ and ‘polarization’. These are measures of the distribution of military capabilities across the civil war system. Fractionalization measures the extent to which the conflict space is characterized by multiple, evenly balanced actors. The more conflict actors and the more equal they are in terms of capabilities, the higher the fractionalization score. Polarization measures the extent to which the conflict space deviates from a 50:50 split. Higher values indicate that the system is closer to this ‘perfect’ bipolarity.

Constructing fractionalization and polarization scores requires information on the number of active rebel groups and their military capabilities. Active rebel groups are those listed in the UCDP/PRIO Armed Conflict Dataset for the relevant conflict year. I have used troop numbers as a proxy for military capability, as these data are available for both government and rebel actors. This is an imperfect measure of capability, but troop numbers are one aspect of the ability to win in open battle, and thus capture an aspect of the concept. Data on troop numbers were sourced primarily from the Cunningham, Gleditsch and Salehyan (2013) Non State Actor Dataset (NSA) and reflects their 'best' estimate of rebel group size. Where these data were missing, the UCDP/PRIO Conflict Encyclopedia was consulted to fill some of these values. Where both these sources do not report numbers on the size of rebel groups, these observations are left as missing. As a robustness check, all missing values on the number of troops connected to active rebel groups were coded as '500' (i.e small) and this did not substantially change the results.

Where a foreign actor is a secondary warring party supporting either the rebels or the government, these troop numbers have been added to the relevant rebel group. In the case where foreign actors intervene after fighting reaches its minimum distance (as when Angola intervened in the Congo in 1997) the relevant group sized are not augmented with the number of foreign troops. Numbers on the extent of foreign support were taken from the UCDP External Actor Dataset and the UCDP Conflict Encyclopedia.

The proportion of the total number of combatant troops was then calculated for each actor by divided the share of each actor by the sum of troop numbers for all actors. These measures are the basis for the construction of the fractionalization and polarization scores, which are calculated with the formulas described in Montalvo and Reynal-Querol (2005).

## **Bibliography**

- Arbetman-Rabinowitz Marina, Ali Fisunoglu, Jacek Kugler, Mark Abdollahian, Kristin Johnson, Kyungkook Kang, Zining Yang (2013) Replication data for: relative political capacity dataset. (<http://hdl.handle.net/1902.1/16845>) Transresearch Consortium, Version 4.
- Buhaug, Halvard; Scott Gates & Päivi Lujala (2009) Geography, rebel capability and the duration of civil conflict. *Journal of Conflict Resolution* 53(4): 544-69.
- Buhaug, Halvard & Scott Gates (2002) The geography of civil war. *The Journal of Peace Research* 39(4): 417-33.

- Carter, David & Curtis Signorino (2010) Back to the future: modeling time-dependence in binary data. *Political Analysis* 18(3): 271-92.
- Cunningham, David E., Kristian Skrede Gleditsch, and Idean Salehyan. 2013. Non-state actors in civil wars: A new dataset. *Conflict Management and Peace Science* 30 (5): 516-531.
- Hallberg, Johan Dittrich. 2012. PRIO Conflict Site 1989-2008: A Geo-Referenced Dataset on Armed Conflict. *Conflict Management and Peace Science* 29 (2): 219-232.
- Fearon, James, and David Laitin. 2003. Ethnicity, Insurgency and Civil War. *American Political Science Review* 97 (1).
- Gleditsch, Kristian S., and Michael D. Ward. 1999. A revised list of independent states since the congress of Vienna. *International Interactions* 25 (4): 393-413.
- Gleditsch, Nils, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Harvard Strand. 2002. Armed Conflict 1946-2001: A New Dataset. *Journal of Peace Research* 39 (5): 615-637.
- Greig, Michael (2014) Rebels at the gates: civil war battle locations, movement, and openings for mediation. *Early View International Studies Quarterly*  
DOI: 10.1111/isqu.12130.
- Heston, Alan, Robert Summers, and Bettina Aten. 2012. "Penn World Table Version 7.1." In, ed. Income and Prices Center for International Comparisons of Production. University of Pennsylvania.
- Kalyvas, Stathis. 2004. The Urban Bias in Research on Civil Wars. *Security Studies* 13 (3): 1-31.
- Montalvo, José G., and Marta Reynal-Querol. 2005. Ethnic Polarization, Potential Conflict and Civil Wars. *American Economic Review* 95 (3): 796-816.
- Moshe, Yegar. 2002. *Between Integration and Secession: The Muslim Communities of the Southern Philippines, Southern Thailand, and Western Burma/Myanmar*. Lanham: Lexington.
- Powell, Jonathan M, and Clayton L Thyne. 2011. Global instances of coups from 1950 to 2010: A new dataset. *Journal of Peace Research* 48 (2): 249-259.
- Raleigh, Clionadh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature. *Journal of Peace Research* 47 (5): 651-660.
- Singer, David. 1987. Reconstructing the Correlates of War Dataset of Material Capabilities, 1816-1985. *International Interactions* 14: 115-132.
- Sundberg, Ralph, Kristine Eck, and Joakim Kreutz. 2012. Introducing the UCDP Non-State Conflict Dataset. *Journal of Peace Research* 49 (2): 351-362.
- Sundberg, Ralph, and Erik Melander. 2013. Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research* 50 (4): 523-532.
- The World Bank (2014) World bank development indicators. edited by The World Bank.